

Юридические исследования

Правильная ссылка на статью:

Бодров Н.Ф., Лебедева А.К. — Понятие дипфейка в российском праве, классификация дипфейков и вопросы их правового регулирования // Юридические исследования. – 2023. – № 11. DOI: 10.25136/2409-7136.2023.11.69014 EDN: DYIHIR URL: https://nbpublish.com/library_read_article.php?id=69014

Понятие дипфейка в российском праве, классификация дипфейков и вопросы их правового регулирования

Бодров Николай Филиппович

ORCID: 0000-0002-9005-3821

кандидат юридических наук

доцент кафедры судебных экспертиз Московского государственного юридического университета имени О.Е. Кутафина (МГЮА)

123242, Россия, Московская область, г. Москва, ул. Садовая-Кудринская, 9, каб. 729

✉ bodrovnf@gmail.com



Лебедева Антонина Константиновна

кандидат юридических наук

доцент, кафедра судебных экспертиз, Московский государственный юридический университет имени О.Е. Кутафина (МГЮА)

125993, Россия, Москва, г. Москва, ул. Садовая-Кудринская, 9

✉ tonya109@yandex.ru



[Статья из рубрики "Трансформация правовых систем"](#)

DOI:

10.25136/2409-7136.2023.11.69014

EDN:

DYIHIR

Дата направления статьи в редакцию:

14-11-2023

Дата публикации:

21-11-2023

Аннотация: В статье авторами рассматриваются вопросы правового регулирования дипфейков в Российской Федерации. Правовое регулирование дипфейков не успевает

за темпами развития технологий искусственного интеллекта. Авторы подчёркивают, что определение дипфейка в действующем законодательстве отсутствует, а существующие в научных работах формулировки крайне противоречивы по своей сути. С учётом темпов развития технологий искусственного интеллекта необходимо законодательно закрепить определение дипфейка. Авторы отмечают, что классификация дипфейков имеет принципиально важное значение для правового регулирования данных технологий. По результатам анализа современных нейросетей предлагается видовая классификация дипфейков. В контексте стремительного развития технологий искусственного интеллекта необходимо должным образом осмыслить нормативное регулирование дипфейков с учетом правовых пробелов, которые сопровождают современный уровень искусственного интеллекта, и потенциальных угроз, которые становятся всё более реальными. Методологическую основу исследования составляют всеобщий диалектический метод, общенаучные (описание, сравнение, обобщение, моделирование и др.) и частнонаучные методы. Основными выводами проведенного исследования является формулирование авторского определения понятия "дипфейк". Дипфейк представляет собой цифровой продукт в виде текста, графики, звука или их сочетания, сгенерированный полностью или частично при помощи нейросетевых технологий, для цели введения в заблуждение или преодоления пользователем систем контроля и управления доступом. Кроме того, авторы предлагают классификацию дипфейков в зависимости от вида генерируемого контента. Дипфейки могут быть созданы в виде текста, графики, звука или их сочетания. Авторы приводят примеры технологий ИИ, использующихся для создания каждого из видов контента. С учетом предлагаемого авторами определения понятия «дипфейк» и с учетом отсутствия правовых механизмов регулирования общественных отношений в сфере использования и распространения дипфейков, обуславливающих развитие цифровой трансформации, важно сформировать механизмы адаптации правовой системы к вызовам, связанным с развитием дипфейк-технологий.

Ключевые слова:

дипфейки, генеративный контент, правовое регулирование, классификация дипфейков, нейросети, синтезированная речь, уголовное право, трансформация права, цифровые технологии, водяные знаки

Статья подготовлена в рамках госзадания «Российская правовая система в реалиях цифровой трансформации общества и государства: адаптация и перспективы реагирования на современные вызовы и угрозы (FSMW-2023-0006)». Регистрационный номер: 1022040700002-6-5.5.1.

Ежедневно в медиапространстве появляются новости об очередных достижениях искусственного интеллекта. Например, на презентации компании Google «Google I/O», состоявшейся 10 мая 2023 года аббревиатура "AI" (сокращение от «artificial intelligence» (искусственный интеллект) была произнесли 146 раз.

С учётом темпов развития технологий искусственного интеллекта (далее – ИИ), доступности продуктов, использующих нейросетевые технологии, возникает необходимость разработки правовых механизмов регулирования результатов применения ИИ, и связанных с ними общественных отношений.

Проблематика искусственного интеллекта и связанных с ними общественных отношений

представляет собой достаточно широкий предмет исследования, который мы рассмотрим с позиции уголовно-правовых наук.

Центральное, по нашим оценкам, место в системе результатов деятельности ИИ место занимают так называемые *deepfake* (дипфейк), обладающие универсальным криминогенным потенциалом.

Нормативно-правовое регулирование дипфейк-технологий и материалов явно не поспевает за темпами технологического развития, но в научной литературе проблематика угроз, связанных с дипфейками, обсуждается с 2019 года [См., например, 3, 7, 10, 12].

Анализ дипфейк-технологий и материалов в силу относительной новизны специфической наукоёмкости в настоящее время в некоторой степени зашел в тупик. Вместо авторских определений дипфейков, учитывающих специфику правового регулирования, в научной литературе часто содержится достаточно примитивная информация из общедоступных источников.

Данный термин используется для обозначения как самой технологии ИИ, так и для номинации самого фейкового контента, что вводит заблуждение как научное сообщество, так и пользователей. Однако те определения дипфейка, которые приводятся в современных научных работах, не раскрывают все грани данного понятия.

Формулируя подходящее определение понятия «дипфейк», мы проанализировали ранее опубликованные в научной литературе определения так как, на наш взгляд, они не в полной мере отражают сущность данного понятия.

Например, в научных работах распространены определения, авторы которых приводят понятие дипфейка без учета возможности создания контента в форме звука:

«Дипфей́ки» — это синтетически произведенный медиаконтент, в котором оригинальный человек (тот, кто изначально находится на изображении) замещается другим человеком» [6, с. 117]. Следует также отметить, далеко не всегда человек «замещается другим человеком», существует возможность замены не только изображения лица или головы, но и трансформация артикуляции и мимики. Кроме того, дипфейк может быть вообще не связан с генерацией изображения человека, мы также можем говорить о генерации, например, конкретного места.

Излишне узким нам представляется подход к анализу дипфейков, когда внимание акцентируется на преднамеренном искажении объектов. Например, в определении:

«Дипфейки - это преднамеренно искаженные аудио, видео или иные файлы с использованием технологии глубокого обучения (определение производное от словосочетания «deep learning» - глубокое обучение, «fake» - подделка), которые изображают что-то вымышленное или ложное, что позволяет злоумышленникам использовать новый и сложный инструмент социальной инженерии» [4, с. 74] автор указывает на преднамеренный характер искажения аудио, видео или иных файлов. Тем не менее, дипфейк – это не какая-то программа для постобработки файлов с целью их изменения, как, например, *voicewriter* (класс программных продуктов для изменения голоса), которые либо посредством постобработки, либо в реальном времени изменяют голос диктора по заранее заложенному в программу алгоритму. Дипфейк создается посредством генерации цифрового продукта на основе обучения нейросети на каком-либо датасете. Это может быть и огромные массивы данных интернета, а также и,

например, фонограммы с записью звучащей речи конкретного лица.

Под «датасетом» в соответствии со ст. 5 "Национальной стратегии развития искусственного интеллекта на период до 2030 года", утвержденной Указом Президента РФ от 10.10.2019 N 490 "О развитии искусственного интеллекта в Российской Федерации" следует понимать:

«д) набор данных - совокупность данных, прошедших предварительную подготовку (обработку) в соответствии с требованиями законодательства Российской Федерации об информации, информационных технологиях и о защите информации и необходимых для разработки программного обеспечения на основе искусственного интеллекта».

С учетом проанализированной специфики следует подойти к определению дипфейка, в первую очередь, как цифрового продукта. Дипфейк в его современном представлении является результатом (продуктом), генерируемым при помощи нейросетевых технологий.

Можно представить и аналоговую форму представления дипфейка, например, опубликованный в периодической печати материал, содержащий изображение или текст, полностью или частично сгенерированные при помощи нейросетевых технологий. Но первоисточником такой информации безусловно будет являться именно цифровой продукт – результат синтеза с использованием нейросетевых технологий.

Дипфейк в цифровой форме может быть представлен разными видами контента или их комбинациями. Поэтому для формулирования определения понятия «Дипфейк» важно, на наш взгляд, определить, в каком виде может быть представлен анализируемый цифровой продукт.

Нам представляется важным проанализировать типы дипфейков по видам генерируемого контента или их сочетаниям.

1. Графика.

Современные технологии позволяют осуществлять синтез графики в виде отдельных изображений или видеофонограмм.

1.1. Изображения.

За последние полгода в интернет-пространстве появилось несколько десятков нейросетей, большинство из которых использует интерфейс текстовых запросов (промпт – от англ. prompt) для синтеза изображений. Text-to-image (генерация изображений по текстовому запросу) модели позволяют генерировать самые различные изображения, включая синтез на основе референтных изображений.

Популярная нейросеть «*Midjourney*», генерирует изображения с возможностью создания графики фотоснимка (с указанием модели камеры и использованной оптики), максимально приближенного к реальному изображению, если задание (промпт) пользователя содержало запрос на это (например, «гиперреалистичный»). Даже по коротким запросам данная нейросеть демонстрирует результаты, способные ввести в заблуждение обывателя. Пользователь может задать качество изображения, с какого устройства это было «снято», расположение объектива, фокусное расстояние и другие параметры реальной фотосъемки. Кроме того, существует команда для генерации изображения по фото. Например, пользователь может загрузить фотографию желаемого лица и сгенерировать практически любую ситуацию с ним.

Безусловно, администрация нейросети разработала некоторые принципы, которые

ограничивают «свободу творчества». Например,

«Не создавайте изображения и не используйте текстовые подсказки, которые по своей сути носят неуважительный, агрессивный или оскорбительный характер. Насилие или домогательства любого рода недопустимы; Никакого контента для взрослых или крови. Пожалуйста, избегайте визуально шокирующего или тревожного контента. Мы автоматически заблокируем ввод некоторых текстов» [\[21\]](#).

Сервис доступа к нейросети «Midjourney» (на платформе Discord) позволяет пользователям выполнять функции администрирования (своего рода социальный контроль пользователями) результатов генерации путем подачи жалоб.

Однако, чтобы создать цифровой продукт, вводящий в заблуждение, не всегда необходимо создавать графическое изображение сцен насилия, жестокости, порнографии и проч. Например, для манипуляции электоратом на выборах может быть достаточно генерации совместного фото представителя лидирующей партии и, предположим, какого-либо военного преступника, сидящих за столом переговоров.

Кроме упомянутой нейросети существуют, например, следующие:

Crayon (ранее Dall-E mini). Нейросеть от разработчиков компании Microsoft, которые подчеркивают, что нейросеть создает крайне реалистичные изображения, что это даже заставляет их переживать, что генерация подобных изображений может привести к непредсказуемым последствиям при создании дипфейков.

Kandinsky 2.2 —нейросеть от разработчиков Сбера при партнёрской поддержке учёных из Института искусственного интеллекта AIRI на объединённом наборе данных Sber AI и SberDevices. Кроме генерации изображения по тексту, данная нейросеть может дорисовывать загруженные фотографии, а также модифицировать их, например, перерисовать существующее изображение в другом стиле или же совместить два изображения.

Шедеврум. Нейросеть от разработчиков Яндекса также имеет собственное мобильное приложение, по существующей информации нейросеть была обучена на 240 млн примеров, которые представляли собой изображения с текстовыми описаниями. Для совершенствования качества обучение нейросети продолжается.

В мае 2023 в Adobe Photoshopпоявился новый инструмент *Generative Fill* (генеративная заливка), на базе ИИ данный инструмент позволяет добавлять, расширять или удалять содержимое изображений без внесения необратимых изменений с помощью простых текстовых запросов. В функционал генеративной заливки включена, например, функция удаления объектов с фотографии, когда пространство вместо объекта (человека, автомобиля следа) автоматически заменяется фоном (без привычных следов обработки инструментами графического редактора).

Характерной особенностью технологических ограничений генеративной заливки на современном этапе является разрешение 1024x1024 (1Мп), но опыт анализа изображений, приобщаемых к материалам дел показывает, что такие возможности заливки более чем достаточны для среднестатистической графики, которую используют в судебном доказывании.

1.2. Видеофонограммы.

Говоря о дипфейках, современные пользователи интернета в первую очередь обычно

представляют именно видеофонограммы. Примеры программ с легкостью можно обнаружить в интернете. Одна из самых известных – *DeepFaceLab*. С помощью данной программы можно заменить лицо в видеофонограмме, изменить возраст диктора на видео, поменять не просто лицо, а форму головы, причёску, голос на фонограмме полностью, а при наличии средних навыков видеомонтажа (например, в *Adobe After Effects* или *Davinci Resolve*) можно трансформировать артикуляцию и мимику лиц на видеофонограмме в дополнение к системе клонирования звучащей речи.

«*DeepFaceLive*» - позволяет заменить лицо как на видеофонограмме, так и в процессе онлайн-общения или в процессе стриминга на веб-камере.

«*Face2Face*» - накладывает мимику управляющего «актёра» на любое другое лицо.

«*Zao Deepfake*» – китайское приложение, которое работает на основе обучаемых нейросетей, позволяет заменять не просто одно лицо фото на другое, но всю мимику и движения лица на видеофонограмме.

Данный перечень можно продолжать и далее, но для общего представления о функционале подобных программ достаточно описания упомянутых нами примеров.

2. Звук.

Если в сентябре 2019 года говорилось, что «имитация» голоса человека крайне трудоемкий и непростой процесс:

«Обучать искусственный интеллект имитировать голос определенного человека дорого и неэффективно, полагает замначальника департамента информационной безопасности Центробанка Артем Сычев: «Применимость таких методов мошенничества крайне низкая. Для этого нужно, чтобы злоумышленник точно знал, что жертва правильно среагирует на этот голос» [\[9\]](#).

То сейчас некоторые сервисы предлагают клонировать вашу речь, обучаясь на небольших фонограммах и предлагая крайне высокое качество аудиосигнала без каких-либо существенных требований к пользователю и доступному ему оборудованию.

Синтез звучащей речи - это технология, позволяющая преобразовывать текст в звучащую речь (TTS-технологии синтеза речи (Text-to-Speech) [см., например, 1, 2, 5].

Говоря о синтезе звучащей речи из текста, нам необходимо отделить его от понятия «клонирование речи» (с англ. voice cloning). Анализируя технологии синтеза звучащей речи важно отметить, что речь идет именно об озвучивании контента голосом «смоделированного» диктора с заданными параметрами. Благодаря технологии глубинного обучения нейросеть может обучаться на фонограммах с голосом и речью миллионов дикторов. Однако, когда перед нейросетью стоит задача генерации звучащей речи конкретного лица, обучение нейросети происходит на фонограммах с образцами голоса и речи индивидуально-определенного диктора. Данный вид синтеза звучащей речи принято называть «клонирование звучащей речи», таким образом, искусственно, с использованием технологий ИИ генерируется звучащая речь реального человека.

С правовой точки зрения важно учесть тот факт, что технология генерация звучащей речи находится на стыке с такой важной индустрией, как биометрическая идентификация по голосу. Биометрическая идентификация/авторизация по голосу уже сейчас используется в качестве самостоятельного инструмента или компонента информационных систем в банковской сфере [\[11\]](#). Так, например в АО «Тинькофф Банк» еще в октябре

2014 года была внедрена подобная технология [\[24\]](#).

Фактически идентификация/аутентификация и клонирование звучащей речи являются конкурирующими технологиями. Правовое регулирование правоотношений в такой сфере должно преследовать не только задачу поспеть за уровнем технологического развития, но и учесть риски отсутствия адекватного и детального регулирования в аспекте обеспечения информационной безопасности.

Рассмотрим некоторые системы синтеза звучащей речи и их возможности.

VALL-E позволяет синтезировать персонализированную речь высокого качества. На основе разработанных алгоритмов и уже осуществленного обучения нейросети *VALL-E* может клонировать голос и речь конкретного диктора, используя, по словам разработчиков, 3-секундные записи, при этом сохраняя эмоции диктора и акустическую обстановку, если это было в запросе пользователя. Данная нейросеть позволяет также создавать текстовый контент, используя генеративные модели ИИ, например, такие как GPT-3 [\[16\]](#).

В мае 2023 года вышла публикация с описанием и примерами генерации звучащей речи «*SoundStorm*» [\[27\]](#). Указанная модель демонстрирует высокие результаты клонирования речи. Авторы публикации, как следует из опубликованных ими материалов, осознают возможности использования их модели для, например, обхода систем биометрической идентификации. Авторы указывают, что с учетом возможного использования модели в противоправных целях, в будущем они планируют изучить подходы к обнаружению синтезированной речи, например, звуковые водяные знаки, чтобы любое потенциальное использование этой технологии строго соответствовало ответственным принципам ИИ, которым придерживаются создатели модели.

ElevenLabs. Как указывают разработчики, они представляют самое реалистичное программное обеспечение для преобразования текста в звучащую речь [\[17\]](#).

Сервис *fakeyou* [\[19\]](#) представляет пользователям возможность клонировать голос любого человека, прежде чем оформить подписку, необходимо ответить на несколько вопросов, для каких целей планируется использовать голос, чей голос будет клонироваться. Для обучения нейросети можно как записать фонограммы через сам сервис, так и прикрепить свои записи высокого качества.

Существуют также сервисы, которые изначально не ставят основной целью синтез и клонирование речи, однако предлагают это пользователям.

Например, *Descript*, данный сервис создан изначально для удобства создателей подкастов и видеороликов, а также их обработки.

Вот некоторые его функции:

- редактор, который одновременно позволяет обрабатывать и видоизменять аудио, текст и видео,
- программа осуществляет распознавание речи и выдает «расшифровку» с текстом ее дословного содержания,
- при редактировании полученного текста изменения происходят также и в аудиоматериале,

– при удалении слов программа автоматически удаляет соответствующие по хронометражу фрагменты видео.

Однако разработчики предлагают также и технологии высококачественного синтеза и клонирования звучащей речи. Lyrebird – это исследовательский отдел ИИ в Descript, создающий новое поколение инструментов для редактирования и синтеза медиаконтента. При помощи возможностей искусственного интеллекта Descript генерирует фразы, точно имитируя особенности голоса человека. Программа может озвучить введенный текст как голосами из имеющейся коллекции, так и собственным голосом пользователя. Программа не работает на русском языке, не распознает и не синтезирует русскую речь.

В заявлении об этике данной компании говорится, что после обучения речевых моделей на ваших образцах голоса (звукозапись начинается только после записи файла устного согласия пользователя), пользователь, как владелец «цифрового голоса», после генерации контролирует то, когда и в каких целях используются результаты синтеза.

Таким образом здесь и в других упомянутых нами примерах отчетливо прослеживается запрос индустрии на создание отраслевых стандартов для борьбы с дезинформацией.

3. Текст

Наибольшие темпы совершенствования на современном этапе развития систем искусственного интеллекта демонстрируют системы, основанные на использовании больших языковых моделей (БЯМ) [\[18\]](#).

Благодаря существенным успехам архитектуры так называемых трансформерных [\[14\]](#) глубоких нейронных сетей широкое распространение получили БЯМ по типу генеративных предобученных трансформеров (GPT).

Наиболее популяризированной среди подобных нейросетей является ChatGPT [\[15\]](#).

Текущая версия - GPT-4, - это языковая модель, разработанная компанией OpenAI, основанная на архитектуре трансформера, а точнее, варианте модели трансформера, известном как трансформер-декодер.

По своей сути ChatGPT - это модель машинного обучения, предназначенная для генерации «человекоподобного» текста, то есть попытке синтеза письменной речи, соответствующей критериям логичности, осмысленности, грамматической правильности и др. Нейросеть обучалась путем машинного обучения, известном как «обучение без наблюдения». При таком подходе модели предоставляется большой корпус текстовых данных, и она учится предсказывать следующее слово в предложении. В частности, модели GPT (Generative Pretrained Transformer) обучаются с помощью варианта бесконтрольного обучения, называемого самоконтрольным или самоконтролируемым обучением.

GPT-4, как и его предшественники, но в большем масштабе, использует огромное количество данных и вычислительных ресурсов для обучения, оперируя в своей работе по разным оценкам порядка 175 миллиардами параметров (с учетом того, что точные данные как и для большинства проприетарных нейросетей в открытом доступе отсутствуют).

С учетом рассматриваемой нами проблематики большие языковые модели представляют практический интерес в плане возможности генерации текстов письменной или

предназначенной для озвучивания речи от имени конкретного лица.

Уже сейчас пользователи имеют возможность задать нейросети такие параметры, чтобы она сгенерировала текст нужной тематики на образцах имеющейся в открытом доступе письменной речи. В качестве образцов могут выступать сообщения из мессенджеров, переписка по электронной почте, публичные записи социальных сетей и т.д. Существует возможность сформулировать задание (prompt) и составить текст на заданную тему в стиле автора (конкретизировать параметры стиля: словарь (лексический запас) автора, особенности употребления синтаксических конструкций, манеру речи и т.д.). Исследование подобного рода запросов уже осуществляется в самостоятельном направлении прикладных исследований – промпт инжиниринге (калька от англ. prompt engineering), одной из основных задач которого является подбор и структурирование модификаций текста запросов (промптов) для решения конкретных задач, например, генерации текста от имени заданного автора.

Сегодня существует техническая возможность генерации текстов на любую тематику. Характерно, что уже сейчас имеются некоторые механизмы этической саморегуляции нейросетевых программных продуктов. Корпорация Open AI, которая является создателями нейросети, указывают, что безопасность пользователя стоит на первом месте, однако в сети существуют инструкции как «обмануть нейросеть» и составить промпты, позволяющие обойти заданные моральные и этические принципы, а также сгенерировать текст без принципа безопасности пользователя.

Среди аналогов ChatGPT в общем доступе находится, например, Bard AI – экспериментальный проект от компании Google, который также как и ChatGPT в виде чат-бота ведёт беседу с пользователем и может генерировать по запросам самые разные тексты.

Важным аспектом рассматриваемой проблематики является общедоступность технологий синтеза текста. Так, например, ресурсы нейросети ChatGPT сейчас доступны фактически каждому пользователю поисковой системы Bing [\[20\]](#), распространяемой корпорацией Microsoft в качестве предустановленной программы-браузера.

С учетом общедоступности можно провести некоторые аналогии с правом интеллектуальной собственности, которое столкнулось с необходимостью правового регулирования общественных отношений в условиях практически неограниченного доступа к информации. Проблематика дипфейк-технологий помимо прочего характеризуется еще и крайне высокой степенью общественной опасности, что, по нашим оценкам, в ближайшем будущем станет одним из первоочередных факторов информационной безопасности.

4. Сочетание Звук+Графика

Сочетание нескольких форм представления, например в виде сочетания сгенерированных звука и графики уже сейчас встречается в форме так называемого «цифрового аватара». Правовому осмыслению данной категории уже посвящен ряд научных работ в юридической литературе [см., например, 8].

Уже несколько сервисов предлагают пользователям создать свой цифровой аватар. Например, сервис «Синтезия» (Synthesia) [\[23\]](#) предлагает пользователям создать свой собственный «AI avatar», то есть аватар, созданный при помощи технологий искусственного интеллекта. Как указывают разработчики, их сервис может создать фотореалистичное изображение реального человека, а также клонировать его голос.

Цифровой аватар создается на основе 15-минутных видео пользователя, а клонирование голоса осуществляется на основе обучения системы на прочтении заранее подготовленного сервисом эталонного текста.

С помощью данного сервиса журналист Джоана Стерн создала свой собственный цифровой аватар, который [\[26\]](#) был использован ею для преодоления механизма системы биометрической верификации по голосу банка, а также введения в заблуждение родственников журналистки.

А при использовании чат-бота ChatGPT возможно обучить нейросеть на образцах письменной или устной речи конкретного лица и синтезировать не только ее голос и внешность, но и содержательную сторону речи. Таким образом, данная форма дипфейка может перейти в следующую: сочетание текста, графики, звука.

Практике уже известны случаи комплексирования различных видов контента, сгенерированного при помощи нейросетей. Так, например, в июне 2023 года в протестантской церкви в Германии была проведена церковная служба, почти полностью сгенерированная при помощи технологий ИИ [\[13\]](#). 40-минутное богослужение, включая проповедь, молитвы и музыку, было создано при помощи ChatGPT, а служба «управлялась» четырьмя разными цифровыми аватарами на экране, двумя молодыми женщинами и двумя молодыми мужчинами.

Ранее упомянутая нами система *DeepFaceLab* позволяет при наличии некоторых навыков в видеомонтаже (например, программы *Adobe After Effects* или *Davinci Resolve*) изменять движение губ лица на видеофонограмме. При использовании систем клонирования звучащей речи на примере, представленном разработчиками, данная нейросеть позволила создателю дипфейка артикулировать нужные фразы, используя внешность известных политических деятелей.

5. Сочетание текста, графики, звука.

Сочетание текста, графики, звука в одном материале с коммерческой точки зрения является даже более популярным видом контента, так как ее развитие тесно связано с запросами разного рода корпоративной индустрии (например создание автоматизированных сервисов новостного вещания, справочных сервисов и проч.).

Несмотря на высокотехнологичность подобного рода контента, на современном этапе его генерация все больше доступна обывателю. Так, например, Китайская компания Tencent Cloud объявила о запуске цифровой платформы для создания людей - *deepfakes-as-a-Service* (DFaaS) [\[25\]](#). По заявлению разработчиков за сравнительно небольшую плату сервис позволяет создавать цифровые копии высокой четкости всех желающих, используя всего три минуты живого видео и 100 произнесенных фраз. Для завершения процесса генерации цифрового аватара потребуется около 24 часов после ввода образцов. То есть технологические возможности позволяют осуществлять подобную генерацию, не требуя от пользователя привлечения собственных вычислительных ресурсов, к тому же позволяя получить результат в короткий срок.

Так, например, создатели сервиса «Spiritme» [\[22\]](#) предлагают пользователям оперативное создание видео с цифровыми аватарами. Для создания цифрового аватара пользователю необходимо записать пятиминутное видео собственной внешности, а в качестве образца может быть использована устная речь на любую тему. Затем сервис генерирует цифровой аватар пользователя, который может произнести любой текст,

который вы подготовите, с достаточной степенью реалистичности а также отображением внешности пользователя, особенностей его голоса и эмоций.

С учётом вышеизложенного криминогенный потенциал такого рода контента крайне сложно переоценить, особенно с учетом степени осведомленности среднестатистического пользователя и в условиях дефицита правовой регламентации технологий и результатов синтеза в отечественной правовой системе.

Анализируя практику по делам, связанным с противоправным распространением дипфейк-контента, и потенциальные угрозы, исходящие от дипфейков, можно сформулировать две главные цели их использования: **введение в заблуждение или преодоление пользователем систем контроля и управления доступом.**

Результаты проведённого анализа сущности дипфейк-контента позволяют нам сформулировать следующее авторское определение дипфейков. **Дипфейк** – цифровой продукт в виде текста, графики, звука или их сочетания, сгенерированный полностью или частично при помощи нейросетевых технологий, для целивведения в заблуждение или преодоления пользователем систем контроля и управления доступом.

В контексте определения понятия «дипфейк» и с учетом отсутствия правовых механизмов регулирования общественных отношений, а также на этапе формирования ряда сквозных цифровых технологий, обуславливающих развитие цифровой трансформации важно сформировать механизмы адаптации правовой системы к вызовам, связанным с развитием дипфейк-технологий. Первоочередными в вопросах правовой регламентации подобного рода технологий и результатов их применения, нам представляются следующие меры:

- нормативное закрепление термина дипфейк и его определение в отечественном законодательстве с учетом технологических возможностей и рисков, вызванных развитием и повсеместным распространением подобных технологий.
- внесение дополнений в нормы уголовного законодательства и законодательства об административных правонарушениях, связанных с применением дипфейк-контента в качестве предмета противоправного (в том числе преступного) посягательства, элемента способа совершения правонарушения (в том числе преступления), обстоятельства,отягчающего ответственность, так как как дипфейк представляет собой более изощренный высокотехнологичный продукт, обладающий принципиально высоким уровнем общественной опасности и трудно распознаваемый на современном уровне развития криминалистической техники.
- внесение дополнений в нормы гражданского законодательства, например, в части защиты чести, достоинства и деловой репутации, а также охраны изображения гражданина, поскольку статьи 152 и 152.1 ГК РФ в должной мере не отражают ставшие уже современными реалии противоправного распространения дипфейк-контента.
- очевидным шагом в регламентации искусственного интеллекта и связанных с ними общественных отношений видится разработка норм профильного законодательства о системах искусственного интеллекта, робототехники, виртуальной реальности, больших данных.

В публично-правовой сфере в первую очередь следует урегулировать вопросы, связанные с ответственностью разработчиков нейросетевых технологий, так как от траектории их развития в первую очередь зависит предмет правового регулирования.

Действенным инструментом контроля за распространением дипфейков могут стать включение в тело файлов, генерируемых нейросетями ватермарок (от англ. watermark) и дополнительной служебной информации в метаданных.

Резюмируя, стоит отметить, что дипфейк-технологии в контексте современной цифровой трансформации являются малоизученными, но одними из наиболее интенсивно развивающихся сфер, с высоким криминогенным потенциалом. Риски, связанные с развитием данной сферы, нам предстоит в полной мере оценить в самое ближайшее время, а от качества нормативно-правовой регламентации сферы применения искусственного интеллекта напрямую зависит объем и характер противоправной деятельности в данной быстроразвивающейся сфере.

Библиография

1. Бодров Н. Ф., Лебедева А. К. Перспективы судебно-экспертного исследования синтезированной звучащей речи / Н. Ф. Бодров, А. К. Лебедева // Законы России: опыт, анализ, практика. 2021. № 3. С. 9-13.
2. Бодров Н. Ф. Судебно-экспертное исследование синтезированной звучащей речи / Н. Ф. Бодров, А. К. Лебедева // Социально-экономическое развитие и качество правовой среды : Сборник докладов VIII Московского юридического форума (XIX Международная научно-практическая конференция): в 5 ч., Москва, 08–10 апреля 2021 года. Том Ч. 4. Москва: Московский государственный юридический университет имени О.Е. Кутафина (МГЮА), 2021. С. 263-266.
3. Даниленко Ю. А. Проблемы расследования отдельных видов киберпреступлений, совершенных с использованием искусственного интеллекта / Ю.А. Даниленко // проблемы получения и использования доказательственной и криминалистически значимой информации. 2019. С. 37-39.
4. Игнатенков Г. К. Технология дипфейк как угроза информационной безопасности / Г. К. Игнатенков // Наука. Исследования. Практика : Сборник избранных статей по материалам Международной научной конференции, Санкт-Петербург, 25 июня 2022 года. Санкт-Петербург: Частное научно-образовательное учреждение дополнительного профессионального образования Гуманитарный национальный исследовательский институт «НАЦРАЗВИТИЕ», 2022. С. 74-77.
5. Лебедева А. К. Технологии голосового синтеза и судебная фоноскопическая экспертиза / А. К. Лебедева // Вестник криминалистики. 2020. № 3(75). С. 55-60.
6. Лужинская Е. Л. Особенности исследования изображений внешнего облика человека, измененного при помощи программных средств / Е. Л. Лужинская, В. А. Чванкин // Вопросы криминологии, криминалистики и судебной экспертизы. 2022. № 2(52). С. 116-121.
7. Овчинников А.В. Распространение Дипфейков В Интернет-Пространстве: Проблемные аспекты правового регулирования / А.В. Овчинников. 2022. Распространение дипфейков в интернет-пространстве. С. 245-255.
8. Рыбаков О.Ю. Человек, право, цифровые технологии: современные направления исследований (обзор Всероссийской научно-практической онлайн-конференции) / Рыбаков О.Ю. // Мониторинг правоприменения. – 2020. – Человек, право, цифровые технологии. № 2 (35). С. 83-87.
9. Сетевое издание «Коммерсантъ» [Электронный ресурс]. URL: <https://www.kommersant.ru/doc/4081979> (дата обращения 30.10.2023).
10. Смирнов А. А. «Глубокие фейки». Сущность и оценка потенциального влияния на национальную безопасность / А.А. Смирнов // Свободная мысль. 2019. № 5 (1677).

С. 63-84.

11. Удаленная идентификация | Банк России [Электронный ресурс]. URL: https://www.cbr.ru/fintech/digital_biometric_id/ (дата обращения: 01.10.2023).
12. Яворский М. А. Deepfake: legal problems and their solution / М.А. Яворский, Т.В. Мавринская // Актуальные проблемы развития правовой системы в цифровую эпоху. 2019. С. 134-138.
13. Apnews [Электронный ресурс]. URL: <https://apnews.com/article/germany-church-protestants-chatgpt-ai-sermon-651f21c24cfb47e3122e987a7263d348> (дата обращения 10.10.2023).
14. Attention is All you Need / A. Vaswani [и др.] // Advances in Neural Information Processing Systems / ред. I. Guyon [и др.]. Curran Associates, Inc., 2017. Т. 30.
15. OpenAI. (2023). ChatGPT (версия от 25 сентября) [большая языковая модель]. <https://chat.openai.com>.
16. Chengyi Wang, Sanyuan Chen, Yu Wu, Ziqiang Zhang Long, Zhou Shujie Liu, Zhuo Chen Yanqing Liu, Huaming Wang, Jinyu Li Lei He Sheng, Zhao Furu Wei (Microsoft). Neural Codec Language Models are Zero-Shot Text to Speech Synthesizers. [Электронный ресурс]. URL: <https://lifeiteng.github.io/valle/index.html> (дата обращения 30.10.2023).
17. ElevenLabs [Электронный ресурс]. URL: <https://beta.elevenlabs.io/speech-synthesis> (дата обращения 28.10.2023).
18. Extracting Training Data from Large Language Models / N. Carlini [и др.] // USENIX Security Symposium. 2020.
19. Fakeyou. [Электронный ресурс]. – URL: <https://fakeyou.com/clone> (дата обращения 30.10.2023).
20. Mehdi Y. Reinventing search with a new AI-powered Microsoft Bing and Edge, your copilot for the web [Электронный ресурс]. URL: <https://blogs.microsoft.com/blog/2023/02/07/reinventing-search-with-a-new-ai-powered-microsoft-bing-and-edge-your-copilot-for-the-web/> (дата обращения: 01.10.2023).
21. Midjourney.su [Электронный ресурс]. URL: <https://midjourney.su/article/usloviya-servisa/?ysclid=libsesc661284902146> (дата обращения: 30.10.2023).
22. SpiritMe [Электронный ресурс]. URL: <https://spiritme.tech> (дата обращения 30.10.2023).
23. Synthesia [Электронный ресурс]. URL: <https://www.synthesia.io/> (дата обращения 30.09.2023).
24. TCS Bank Is First Among Russian Banks to Introduce Voice Authentication System for Its Call Centre — Tinkoff news [Электронный ресурс]. URL: <https://www.tinkoff.ru/about/news/21102014-tcs-introduce-voice-authentication-system/> (дата обращения: 14.10.2023).
25. Tencent Cloud [Электронный ресурс]. URL: <https://www.tencentcloud.com/> (дата обращения: 01.10.2023).
26. The Wall Street Journal [Электронный ресурс]. URL: <https://www.wsj.com/articles/i-cloned-myself-with-ai-she-fooled-my-bank-and-my-family-356bd1a3> (дата обращения 30.10.2023).
27. Zalán Borsos, Matt Sharifi, Damien Vincent, Eugene Kharitonov, Neil Zeghidour, Marco Tagliasacchi. SoundStorm: Efficient Parallel Audio Generation. [Электронный ресурс]. URL: <https://google-research.github.io/seanet/soundstorm/examples/> (дата обращения 30.10.2023).

Результаты процедуры рецензирования статьи

В связи с политикой двойного слепого рецензирования личность рецензента не раскрывается.

Со списком рецензентов издательства можно ознакомиться [здесь](#).

Предметом исследования в представленной на рецензирование статье являются, как это следует из ее наименования, понятие дипфейка в российском праве, классификация дипфейков и вопросы их правового регулирования. Заявленные границы исследования обозначены автором ("Проблематика искусственного интеллекта и связанных с ними общественных отношений представляет собой достаточно широкий предмет исследования, который мы рассмотрим с позиции уголовно-правовых наук"), но фактически содержание статьи выходит за узкие "уголовно-правовые" рамки и соответствует ее "общеотраслевому" наименованию. Однако название работы несколько громоздкое и больше подходит для учебной, а не научной работы. Поэтому рекомендуется изменить его на "Феномен дипфейков: актуальные вопросы правового регулирования" или т.п.).

Методология исследования в тексте статьи не раскрывается, но очевидно, что ученым использовались всеобщий диалектический, логический, статистический, формально-юридический, сравнительно-правовой методы исследования.

Актуальность избранной автором темы исследования несомненна и обоснована им следующим образом: "Ежедневно в медиапространстве появляются новости об очередных достижениях искусственного интеллекта. Например, на презентации компании Google «Google I/O», состоявшейся 10 мая 2023 года аббревиатура «AI» (сокращение от «artificial intelligence» (искусственный интеллект) была произнесли 146 раз. С учётом темпов развития технологий искусственного интеллекта (далее – ИИ), доступности продуктов, использующих нейросетевые технологии, возникает необходимость разработки правовых механизмов регулирования результатов применения ИИ, и связанных с ними общественных отношений"; "Центральное, по нашим оценкам, место в системе результатов деятельности ИИ место занимают так называемые deerfake (дипфейк), обладающие универсальным криминогенным потенциалом". Ученым раскрыта степень изученности рассматриваемых в статье проблем: "Нормативно-правовое регулирование дипфейк-технологий и материалов явно не поспевает за темпами технологического развития, но в научной литературе проблематика угроз, связанных с дипфейками, обсуждается с 2019 года [См., например, 3, 7, 10, 12]. Анализ дипфейк-технологий и материалов в силу относительной новизны специфической наукоёмкости в настоящее время в некоторой степени зашел в тупик. Вместо авторских определений дипфейков, учитывающих специфику правового регулирования, в научной литературе часто содержится достаточно примитивная информация из общедоступных источников".

Научная новизна работы проявляется в формировании автором современного целостного представления о феномене дипфейка и его типах. Непосредственный интерес представляет ряд заключений ученого: "... криминогенный потенциал такого рода контента крайне сложно переоценить, особенно с учетом степени осведомленности среднестатистического пользователя и в условиях дефицита правовой регламентации технологий и результатов синтеза в отечественной правовой системе. Анализируя практику по делам, связанным с противоправным распространением дипфейк-контента, и потенциальные угрозы, исходящие от дипфейков, можно сформулировать две главные цели их использования: введение в заблуждение или преодоление пользователем систем контроля и управления доступом. Результаты проведенного анализа сущности

дипфейк-контента позволяют нам сформулировать следующее авторское определение дипфейков. Дипфейк – цифровой продукт в виде текста, графики, звука или их сочетания, сгенерированный полностью или частично при помощи нейросетевых технологий, для цели введения в заблуждение или преодоления пользователем систем контроля и управления доступом". Автор вносит ряд ценных с теоретической и практической точек зрения предложений: "... нормативное закрепление термина дипфейк и его определение в отечественном законодательстве с учетом технологических возможностей и рисков, вызванных развитием и повсеместным распространением подобных технологий; внесение дополнений в нормы уголовного законодательства и законодательства об административных правонарушениях, связанных с применением дипфейк-контента в качестве предмета противоправного (в том числе преступного) посягательства, элемента способа совершения правонарушения (в том числе преступления), обстоятельства,отягчающего ответственность, так как как дипфейк представляет собой более изощренный высокотехнологичный продукт, обладающий принципиально высоким уровнем общественной опасности и трудно распознаваемый на современном уровне развития криминалистической техники; внесение дополнений в нормы гражданского законодательства, например, в части защиты чести, достоинства и деловой репутации, а также охраны изображения гражданина, поскольку статьи 152 и 152.1 ГК РФ в должной мере не отражают ставшие уже современными реалии противоправного распространения дипфейк-контента; очевидным шагом в регламентации искусственного интеллекта и связанных с ними общественных отношений видится разработка норм профильного законодательства о системах искусственного интеллекта, робототехники, виртуальной реальности, больших данных". Таким образом, статья вносит определенный вклад в развитие отечественной правовой науки и, безусловно, заслуживает внимания читательской аудитории.

Научный стиль исследования выдержан автором в полной мере.

Структура работы вполне логична. Во вводной части статьи ученый обосновывает актуальность избранной им темы исследования. В основной части работы автор осуществляет критический анализ основных теоретических подходов к пониманию сущности понятия "дипфейк", рассматривает типы дипфейков по видам генерируемого контента или их сочетаниям, предлагает оригинальную дефиницию исследуемого понятия, а также комплекс мер по совершенствованию соответствующего правового регулирования. В заключительной части статьи содержатся выводы по результатам проведенного исследования.

Содержание статьи полностью соответствует ее наименованию и не вызывает особых нареканий.

Библиография исследования представлена 27 источниками (научными статьями, аналитическими и публицистическими материалами), в том числе на английском языке. С формальной и фактической точек зрения этого вполне достаточно. Характер и количество использованных при написании статьи источников позволили автору раскрыть тему исследования с необходимой глубиной и полнотой.

Апелляция к оппонентам имеется, как общая, так и частная (Е. Л. Лужинская, Г. К. Игнатенков и др.), и вполне достаточна. научная дискуссия ведется автором корректно. Положения работы обоснованы в необходимой степени.

Выводы по результатам проведенного исследования имеются ("Анализируя практику по делам, связанным с противоправным распространением дипфейк-контента, и потенциальные угрозы, исходящие от дипфейков, можно сформулировать две главные цели их использования: введение в заблуждение или преодоление пользователем систем контроля и управления доступом. Дипфейк – цифровой продукт в виде текста, графики, звука или их сочетания, сгенерированный полностью или частично при помощи

нейросетевых технологий, для целивведения в заблуждение или преодоления пользователем систем контроля и управления доступом. ... Первоочередными в вопросах правовой регламентации подобного рода технологий и результатов их применения, нам представляются следующие меры: - нормативное закрепление термина дипфейк и его определение в отечественном законодательстве с учетом технологических возможностей и рисков, вызванных развитием и повсеместным распространением подобных технологий. - внесение дополнений в нормы уголовного законодательства и законодательства об административных правонарушениях, связанных с применением дипфейк-контента в качестве предмета противоправного (в том числе преступного) посягательства, элемента способа совершения правонарушения (в том числе преступления), обстоятельства, отягчающего ответственность, так как как дипфейк представляет собой более изощренный высокотехнологичный продукт, обладающий принципиально высоким уровнем общественной опасности и трудно распознаваемый на современном уровне развития криминалистической техники. - внесение дополнений в нормы гражданского законодательства, например, в части защиты чести, достоинства и деловой репутации, а также охраны изображения гражданина, поскольку статьи 152 и 152.1 ГК РФ в должной мере не отражают ставшие уже современными реалии противоправного распространения дипфейк-контента. - очевидным шагом в регламентации искусственного интеллекта и связанных с ними общественных отношений видится разработка норм профильного законодательства о системах искусственного интеллекта, робототехники, виртуальной реальности, больших данных. В публично-правовой сфере в первую очередь следует урегулировать вопросы, связанные с ответственностью разработчиков нейросетевых технологий, так как от траектории их развития в первую очередь зависит предмет правового регулирования. Действенным инструментом контроля за распространением дипфейков могут стать включение в тело файлов, генерируемых нейросетями ватермарок (от англ. watermark) и дополнительной служебной информации в метаданных"), они четкие, конкретные, обладают свойствами научной новизны, достоверности и обоснованности. Таким образом, выводы заслуживают внимания читательской аудитории.

В статье встречаются опечатки, поэтому она нуждается в дополнительном вычитывании. Так, автор пишет: "Например, на презентации компании Google «Google I/O», состоявшейся 10 мая 2023 года аббревиатура "AI" (сокращение от «artificial intelligence» (искусственный интеллект) была произнесли 146 раз" - "была произнесена". Ученый отмечает: "То сейчас некоторые сервисы предлагают клонировать вашу речь, обучаясь на небольших фонограммах и предлагая крайне высокое качество аудиосигнала без каких-либо существенных требований к пользователю и доступному ему оборудованию" - опечатка в начале предложения. Автор указывает: "С правовой точки зрения важно учесть тот факт, что технология генерация звучащей речи находится на стыке с такой важной индустрией, как биометрическая идентификация по голосу" - "технология генерации", а не "технология генерация".

Интерес читательской аудитории к представленной на рецензирование статье может быть проявлен прежде всего со стороны специалистов в сфере теории государства и права, информационного права, гражданского права, административного права, уголовного права при условии ее небольшой доработки: уточнении наименования работы, раскрытии методологии исследования, устранении нарушений в оформлении статьи (опечаток).